Origin of Peer Influence in Social Networks

Flávio L. Pinheiro,[1,2,3] Marta D. Santos,[4] Francisco C. Santos,[3,1] and Jorge M. Pacheco[5,6,1]
[1]*ATP Group, CMAF, Instituto para a Investigação Interdisciplinar, P-1649-003 Lisboa, Portugal*
[2]*Centro de Física da Universidade do Minho, 4710-057 Braga, Portugal*
[3]*INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, 2744-016 Porto Salvo, Portugal*
[4]*Departamento de Física and I3N, Universidade de Aveiro, 3810-193 Aveiro, Portugal*
[5]*Departamento de Matemática e Aplicações da Universidade do Minho, 4710-057 Braga, Portugal*
[6]*Centro de Biologia Molecular e Ambiental da Universidade do Minho, 4710-057 Braga, Portugal*
(Received 6 May 2012; revised manuscript received 3 October 2013; published 6 March 2014)

Social networks pervade our everyday lives: we interact, influence, and are influenced by our friends and acquaintances. With the advent of the World Wide Web, large amounts of data on social networks have become available, allowing the quantitative analysis of the distribution of information on them, including behavioral traits and fads. Recent studies of correlations among members of a social network, who exhibit the same trait, have shown that individuals influence not only their direct contacts but also friends' friends, up to a network distance extending beyond their closest peers. Here, we show how such patterns of correlations between peers emerge in networked populations. We use standard models (yet reflecting intrinsically different mechanisms) of information spreading to argue that empirically observed patterns of correlation among peers emerge naturally from a wide range of dynamics, being essentially independent of the type of information, on how it spreads, and even on the class of underlying network that interconnects individuals. Finally, we show that the sparser and clustered the network, the more far reaching the influence of each individual will be.

DOI: 10.1103/PhysRevLett.112.098702 PACS numbers: 89.65.Ef, 87.23.Kg, 89.75.Fb, 89.75.Hc

Human societies are embedded in complex social networks on which information flow—associated with traits such as emotions, behaviors, ideas, or fads—is ubiquitous [1–12]. What determines the patterns observed has become extremely valuable, with applications extending to all areas of human activity. Several studies have focused on the role played by social networks on the spread of information between individuals, by making use of email and blog databases and online social networks such as Twitter [11,12] and Facebook [6,7]. Such empirical studies have shown how social networks affect the propagation of health issues [13,14], ideas [15], criminal behavior [16,17], economic decisions [18,19], school achievement [20], and cooperation [21,22], among other human traits.

In what concerns the correlation patterns observed, Fowler and Christakis [22–25] recently proposed a *three degree of influence* rule based on the statistical analysis of the Framingham Heart Study database, from which a social network was inferred. Correlations among individuals were analyzed for traits as diverse as smoking habits, alcohol consumption, loneliness, obesity, cooperation, or happiness. These correlations reflected the relative increase in probability—when compared with a random arrangement—that two individuals share the same trait as a function of the network distance, defined as the smallest number of links connecting those individuals in the network [see Fig. 1(a)]. They found that similar and nontrivial correlations emerge from distinct traits and persist in the period studied, suggesting the validity of a three degrees of influence on social networks. In other words, not only our "friends" but also our friends' friends together with their friends exhibit a positive correlation of traits. More recently, analysis of cooperation on social networks of hunter-gatherers revealed a degree of influence of 2 [26].

Here, we investigate the degree of peer influence that emerges from different dynamical processes representative of a plethora of phenomena occurring in networked populations—the spread of cooperative strategies, opinions, and diseases. Individuals are assigned to nodes of a complex network, whereas links between them represent interactions. We show that, for each network class considered, different processes often lead to the same degrees of influence, suggesting that peer influence is insensitive to the process at stake. On the other hand, we find that simple topological properties of the underlying networks, such as the average connectivity ($\langle k \rangle$) and the clustering coefficient, ultimately determine the number of degrees of influence observed, which systematically falls between 3 and 2, in agreement with the results stemming from empirical analyses of correlations in present [22–25] and past [26] social networks.

We start by studying the evolutionary dynamics of cooperation [5,27], modeled as peer-to-peer interactions by means of the famous Prisoner's Dilemma (PD) metaphor [27]. The word dilemma in the PD stems from the decision conflict that occurs when 2 individuals must simultaneously decide whether to Cooperate (C) or to Defect (D) towards the other. The game returns $R = 1$ for mutual

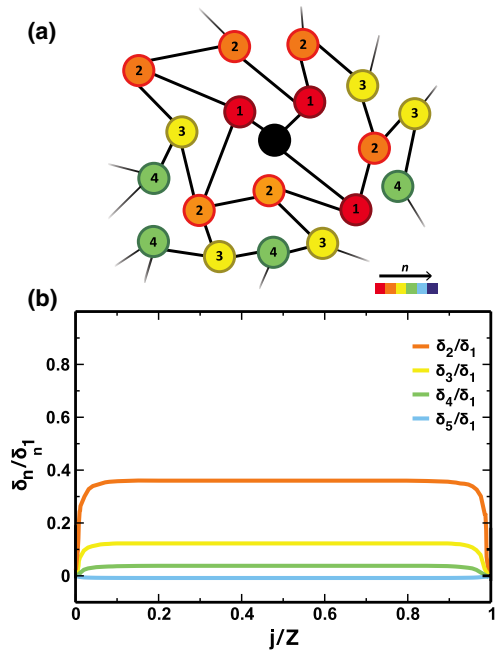

FIG. 1 (color online). (a) Network distance of a given node to the focal individual (black circle) defined as the smallest number of network links that separate the two (given by the number inside each circle). (b) Examples of correlation patterns as a function of j/Z (where j is the number of cooperators present in a population of size Z), emerging from the evolutionary PD game in homogeneous small world networks of size $Z = 10^3$ and degree $\langle k \rangle = 4$; the ratio $\delta_n(j/Z)/\delta_1(j/Z)$ provides an adequate normalization that renders the correlation patterns approximately constant for most values of j/Z. Similar patterns are obtained both for VM and SIR dynamics.

cooperation, $P = 0$ for mutual defection, $S = -\lambda$ when playing C against a D, and $T = 1 + \lambda$ when playing D against a C ($\lambda > 0$ measures both the temptation to defect and the fear of being cheated [28]). The ranking $T > R > P > S$ implies that maximizing one's own payoff leads each individual to choose D, irrespective of the decision of the other, such that the outcome will be mutual defection (pure Nash equilibrium for $\lambda > 0$). This outcome, however, is not the best for the pair, as mutual cooperation would lead to a better outcome for both of them. In the evolutionary version, the accumulated return from the interactions with all neighbors is interpreted as a measure of success (or "fitness"), such that some strategies (C or D) may become more attractive than others. We assume that individuals revise their behavior based on the perceived success of others: an individual A imitates a randomly chosen neighbor B with probability $p = [1 + e^{-\beta(f_B - f_A)}]^{-1}$, where f_A (f_B) stands for the fitness of A (B) and β denotes the intensity of selection [29]. In the mean-field limit, in which everyone is equally likely to interact with anyone else (also known as well-mixed population approximation), this dilemma inexorably condemns cooperation to extinction [27], a fate which may change when individuals are embedded in a social network represented by means of a graph, in which structural diversity is ubiquitous [1–3,5,7–10,22–26,28,30–35].

In practice, however, the (population) dynamics of peer influence does not need to be fitness-driven or constrained by any type of social dilemma. To investigate opinion dynamics, we adopt a variant of one of the simplest and most commonly used models of opinion formation—the Voter Model (VM)—in which an individual can be influenced by a randomly chosen neighbor to adopt the opinion of the latter with probability $p = 1.0$ [10,36–38]. We confirmed that lower values of p do not change the results discussed below.

Moreover, information transmission has been sometimes regarded as contagious [15,39], similar to the propagation of infectious diseases [10]. Hence, we also study a widely used model of disease spreading, the Susceptible-Infected-Recovered model (SIR) [40], in which individuals can be in a susceptible (S), infected (I), or recovered (R) epidemiological state [10,40]. An infected individual can either infect a susceptible neighbor at an infection rate α or recover and become immune at a recovery rate γ.

Given the structural diversity of social networks [2,3,30,34] in which some individuals interact and/or are taken as role models more and more often than others, we investigate the role played by the distribution of the number of first neighbors of each individual (the degree distribution [2]) and clustering [2,35] (a measure of the number of individuals' neighbors who are also neighbors of each other) in the emerging patterns of correlations. To this end, we employ four network classes [30,41] with increasing variance of the degree distribution and similar low levels of clustering: homogeneous and heterogeneous small-world networks and exponential and scale-free networks [2,3,30]. Homogeneous small-world (HoSW) networks were obtained by repeatedly swapping the ends of pairs of randomly chosen links of a regular ring. Scale-free networks were obtained combining growth and preferential attachment, following the model proposed by Barabási and Albert (BA) [2]. Exponential networks (EXP) were obtained by adopting the same algorithm, with preferential attachment replaced by random attachment [2]. Heterogeneous small-world (HeSW) networks were built adopting the limit $p = 1$ of the Watts-Strogatz model [35], in which all links are rewired. Results reported in the main text refer to networks with $Z = 10^3$ (see Fig. 2) nodes, of the same order of magnitude of those investigated in Refs. [22–25] and for which finite-size effects are non-negligible. In the Supplemental Material [41], we investigate in detail the behavior of correlations as a function of the network size.

Let j be the number of individuals carrying one of the traits in a population of size Z. For each dynamical process and each j/Z we determine the propensity $\delta_n(j/Z)$ that 2 individuals at a distance n self-organize in the same trait

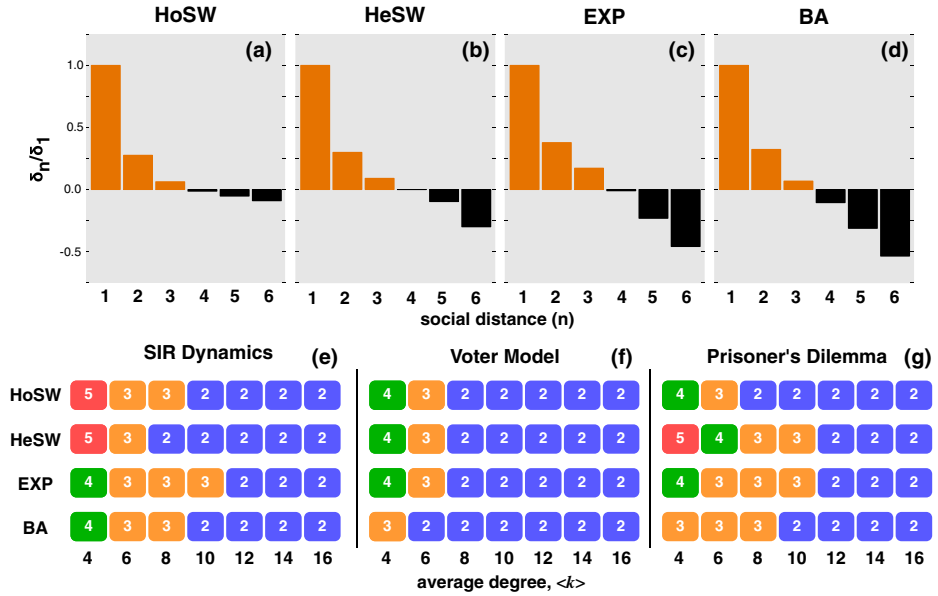

FIG. 2 (color online). Peer influence in social networks. Upper panels: δ_n/δ_1 for recovered and infected individuals in the SIR epidemic model taking place in (a) homogeneous small world networks, (b) heterogeneous small world networks, (c) exponential networks, and (d) scale-free networks. Lower panels: Critical degree of peer influence n_c (defined as the largest network distance n for which $\delta_n > 0$) for different values of average degree $\langle k \rangle$ and network classes shown in the top panels, for the following traits and processes (see main text for details). (e) Recovered and infected individuals in the SIR epidemic model (as above), (f) individuals with the same opinion in the VM, and (g) cooperators in the evolutionary PD game. All panels correspond to networks of size $Z = 10^3$, and results in the upper panels were obtained for networks with an average degree $\langle k \rangle = 6$. For larger network (and population) sizes, the negative values of the peer-influence correlation amplitudes that one obtains above n_c (see upper panels) will slowly approach 0, highlighting the importance of finite-size effects in evaluating peer-to-peer processes. The reason for the negative correlation values for $n > n_c$ in finite (and small) networks is clear: at a network distance n_c, most of the individuals with the same trait have been already sampled (see Fig. 2 in the Supplemental Material [41]); thus, for larger network distances, negative correlations will inevitably build up. At any rate, it is possible to clearly identify two regions of influence separated by n_c (see the Supplemental Material [41] for additional details).

relative to a random distribution of traits [23,24]. This quantity can be written as $\delta_n(j/Z) = \epsilon_n(j/Z)/\epsilon_n^{\text{rand}}(j/Z) - 1$, where $\epsilon_n(j/Z)$ is the average probability that a node shares the same trait with nodes located at a network distance n and $\epsilon_n^{\text{rand}}(j/Z)$ is the same quantity associated with a random distribution of traits, given by j/Z. The network distance (n) between individuals corresponds to the smallest number of links separating two nodes of a complex network, as depicted in Fig. 1(a). Additionally, we define n_c as the largest number of links (network distance) n for which δ_n remains positive ($n_c = 3$ in Refs. [22–25] and $n_c = 2$ in Ref. [26]).

To obtain $\epsilon_n(j/Z)$ we compute, for each network configuration, the average fraction of nodes that exhibit the same trait at a distance n. Hence, for each dynamical process, $\epsilon_n(j/Z)$ results from averaging over 10^6 independent network configurations. For the PD and VM processes, simulations where carried out starting from random configurations of traits; configurations included in the average were extracted after a transient of 10^3 generations (1 generation $= Z$ iterations). For the SIR process, each simulation was started with a single I in a population of S; α and γ where defined such that the population would often reach a state with no Is left (Supplemental Material [41]).

The upper panels of Fig. 2 illustrate the normalized correlation values δ_n/δ_1, which as shown in Fig. 1(b) are approximately independent from j/Z, obtained for the SIR dynamics in HoSW, HeSW, EXP, and BA networks. We observe that, regardless of the network topology, $n_c = 3$ as in Refs. [22–26]. Similar trends are also obtained for the two other dynamical processes introduced above (Supplemental Material [41]). This is shown in the lower panels of Fig. 2, which also show how the critical network distance (n_c) stabilizes at $n_c = 2$ with increasing network connectivity (e.g., $\langle k \rangle$), independently of the network structure, cluster coefficient, and dynamical process (see also the Supplemental Material [41]). On the contrary, deviations from this universal behavior of n_c are obtained only whenever networks become very sparse, associated with the smallest values of the average connectivity. Counterintuitively, the sparser the underlying network, the more far reaching the influence of each individual will be, extending beyond the most pervasive value of 2. Needless to say, the absolute values of the correlations, depicted in the

upper panels, will depend on the specific model parameters but not the final value of n_c.

Besides being heterogeneous, social networks often exhibit sizable levels of clustering [2], contrary to the negligible values that characterize the ones utilized in Fig. 2. To evaluate the impact of this property, we generate in the Supplemental Material [41] networks with arbitrary clustering for each network class [41,42]. We show how n_c remains limited between 2 and 4 for each of the three heterogeneous networks under study, irrespectively of their clustering coefficient and average degree. Nonetheless, we observe that increasing levels of clustering act to enlarge n_c, mostly whenever networks are very sparse, thus sizeably increasing their average path length (Supplemental Material [41]).

Overall, our results suggest that the extent of peer influence emerges as a natural outcome of dynamical processes on structured populations, being pervasive in a wide range of phenomena occurring in social networks. Despite the importance of social networks in defining the paths and ends of the dynamical processes they support, showing how important it is to address and understand population dynamics from a complex networks perspective [5,10,32,33,37], the patterns of peer influence they exhibit are surprisingly independent of their structure. On the other hand, when networks are very sparse, different network properties may contribute to enlarge the sphere of influence of each individual. Our results also show how networks naturally entangle individuals into interactions of many-body nature: Indeed, social networks effectively extend, in nontrivial ways, the dyadic interactions we started from. The fact that the network distance between any two individuals in social networks is small [2,3] and comparable to n_c further enhances the significance of the present results, as they stress how our individual actions may have wide scopes and counterintuitive repercussions.

Financial support by FEDER through POFC—COMPETE and by FCT-Portugal is gratefully acknowledged through Grants No. SFRH/BD/77389/2011, No. SFRH/BPD/90936/2012, No. PTDC/MAT/122897/2010, No. EXPL/EEI-SII/2556/2013, No. PEst-OE/EEI/LA0021/2013, and No. PEst-OE/BIA/UI4050/2014.

[1] A. L. Lloyd and R. M. May, Science **292**, 1316 (2001).
[2] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] S. N. Dorogotsev and J. F. F. Mendes, *Evolution on Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, New York, 2003).
[4] M. Rosvall and K. Sneppen, Phys. Rev. Lett. **91** 178701 (2003).
[5] G. Szabó and G. Fáth, Phys. Rep. **446**, 97 (2007).
[6] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, Soc. Networks **30**, 330 (2008).
[7] J. P. Onnela and F. Reed-Tsochas, Proc. Natl. Acad. Sci. U.S.A. **107**, 18 375 (2010).
[8] D. Liben-Nowell and J. Kleinberg, Proc. Natl. Acad. Sci. U.S.A. **105**, 4633 (2008).
[9] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Phys. Rep. **469**, 93 (2008).
[10] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
[11] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummad, in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC* (2010).
[12] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, in *Proceedings of the Fourth International Conference on Web Search and Data Mining, Hong-Kong* (2011), p. 65.
[13] J. G. Trogdon, J. Nonnemaker, and J. Pais, J. Health Econ. **27**, 1388 (2008).
[14] H. Nair, P. Manchanda, and T. Bhatia, J. Market. Res. **47**, 883 (2010).
[15] L. Bettencourt, A. Cintron-Arias, D. Kaiser, and C. Castillo-Chavez, Physica (Amsterdam) **364A**, 513 (2006).
[16] E. L. Glaeser, B. Sacerdore, and J. A. Scheinkman, Q. J. Econ. **111**, 507 (1996).
[17] A. Calvó-Armengol and Y. Zenou, Int. Econ. Rev. **45**, 939 (2004).
[18] A. Calvó-Armengol, J. Econ. Theory **115**, 191 (2004).
[19] L. Cohen, A. Fazzini, and C. Malloy, J. Polit. Econ., **116**, 951 (2008).
[20] B. Sacerdote, Q. J. Econ. **116**, 681 (2001).
[21] P. Brañas-Garza, R. Cobo-Reyes, M. Paz Espinosa, N. Jiménez, J. Kovářík, and G. Ponti, Games Econ. Behav. **69**, 249 (2010).
[22] J. Fowler and N. Christakis, Proc. Natl. Acad. Sci. U.S.A. **107**, 5334 (2010).
[23] N. Christakis, and J. Fowler, N. Engl. J. Med. **358**, 2249 (2008).
[24] N. Christakis and J. Fowler, N. Engl. J. Med. **357**, 370 (2007).
[25] J. Fowler and N. Christakis, Br. Med. J. **337**, a2338 (2008).
[26] C. L. Apicella, F. W. Marlowe, J. H. Fowler, and N. A. Christakis, Nature (London) **481**, 497 (2012).
[27] K. Sigmund, *The Calculus of Selfishness* (Princeton University Press, Princeton, NJ, 2010).
[28] F. C. Santos, J. M. Pacheco, and T. Lenaerts, Proc. Natl. Acad. Sci. U.S.A. **103**, 3490 (2006).
[29] G. Szabó and C. Tőke, Phys. Rev. E **58**, 69 (1998).
[30] L. A. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[31] F. C. Santos and J. M. Pacheco, Phys. Rev. Lett. **95**, 098104 (2005).
[32] M. Perc and A. Szolnoki, BioSystems **99**, 109 (2010).
[33] F. L. Pinheiro, J. M. Pacheco, and F. C. Santos, PLoS One **7**, e32114 (2012).
[34] J. P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A.-L. Barabasi, Proc. Natl. Acad. Sci. U.S.A. **104**, 7332 (2007).
[35] D. Watts and S. Strogatz, Nature (London) **393**, 440 (1998).
[36] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes* (Springer–Verlag, Berlin, 1999).
[37] T. Antal, S. Redner, and V. Sood, Phys. Rev. Lett. **96**, 188104 (2006).

[38] V. Sood, T. Antal, and S. Redner, Phys. Rev. E **77** 041121 (2008).
[39] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, *Emotional Contagion* (Cambridge University Press, Cambridge, England, 1994).
[40] W. O. Kermack and A. G. McKendrick, Proc. R. Soc. A **115**, 700 (1927).
[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.112.098702 for additional details of the models and methods used, and for results obtained for larger network sizes.
[42] S. Bansal, S. Khandelwal, and L. A. Meyers, BMC Bioinf. **10**, 405 (2009).